2013

Spatialized Audio and Landmarks in Team Navigation

Andrew Hampton
Wright State University

SPATIALIZED AUDIO AND LANDMARKS IN TEAM NAVIGATION

A thesis submitted in partial fulfillment of the
requirements for the degree of
Master of Science

By

ANDREW HAMPTON

B.S., University of Central Florida, 2009

2013

Wright State University

WRIGHT STATE UNIVERSITY

GRADUATE SCHOOL

MAY 3, 2013

I HEREBY RECOMMEND THAT THE THESIS PREPARED UNDER MY SUPERVISION BY Andrew Hampton ENTITLED Spatialized Audio and Landmarks in Team Navigation BE ACCEPTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF Master of Science.

Valerie L. Shalin, Ph.D.
Thesis Director

Scott N.J. Watamaniuk, Ph.D.
Graduate Program Director

John M. Flach, Ph.D.
Chair, Department of Psychology

Committee on Final Examination

Valerie L. Shalin, Ph.D.

Robert H. Gilkey, Ph.D.

Herbert A. Colle, Ph.D.

R. William Ayres, Ph.D.
Interim Dean, Graduate School

ABSTRACT

Hampton, Andrew. M.S. Department of Psychology, Wright State University, 2013. Spatialized Audio and Landmarks in Team Navigation.

Using data collected from a prior study that established the benefit of spatialized audio on team navigation, the current paper examines the underlying mechanisms by which that benefit arose. With linguistic measures extracted from trial transcripts, I study emergent patterns of conversation for four dyads as they attempt to rendezvous in an immersive virtual environment. Spatialized audio is compared to landmarks, traditionally viewed as integral to navigation tasks, on the basis of coordination and strategy. Analyses reveal that spatialized audio creates a decreased need to speak overall. Paired with the performance advantage, this creates a more linguistically efficient task structure. Spatialized audio may change the perspective of participants, giving them a more comprehensive view of their environment. Furthermore, the absence of changes in coordination measures due to landmark manipulation introduces the idea that landmarks are a purely cognitive construct, not necessarily defined in real-time (as opposed to planned or recalled) navigation tasks.

Keywords: spatialized audio, landmarks, navigation, coordination, language

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

I. INTRODUCTION

This paper examines communication during a team navigation task. From taking a road trip to a military extraction behind enemy lines, difficulty navigating can result in negative consequences ranging from a simple wrong turn to the loss of personnel. The conventional approach to communication within a team conducting a navigation task relies heavily on verbal references to visible landmarks. However, audible cues, such as the sound of river rapids, can perform a similar function. Such cues seem to emanate from a particular location in space.

Spatialized audio technology that provides digitally transmitted sound provides the capability to designate auditory cues at will. Whereas landmarks require specification with a symbol (often a verbal label) to communicate to others, spatialized audio does not. According to Peirce (1991), symbols bear an arbitrary relationship to the world, and depend on convention and knowledge. In contrast, spatialized audio cues do not require representation with symbols. In the natural environment, sound indexes its source in an inherently non-arbitrary way, and therefore requires less interpretation.

I refer to the use of symbols during communication in team navigation as symbolic cuing, in recognition of the role of higher-order thought in interpretation. Spatialized audio communications may function more like auditory cues in the natural environment, producing perceptual data without the need for symbolic processing.

A team navigation task provides an optimal domain to investigate the function of symbolic and auditory cues, using language as a measure. Past research has addressed language measures across a range of small scale spatial tasks, generally puzzles and assembly (e.g., Bangerter & Clark, 2003; Ou, Oh, Fussell, Blum, & Yang 2008). This research provides a foundation of expectations for verbal behavior in the present task, but the previous tasks did not include a role for sound source as a carrier of information. In the large scale, dynamic environments examined here, spatio-temporal coordinated activities require joint navigation to achieve a shared goal. Speech functions as both a source of symbolic content for labeling landmarks, and, in the case of spatialized audio, as a sound that indexes the speaker's location. In the latter case, we can expect spatialized audio to affect the need for symbolic language.

In the following sections I note the effects of cognitive versus perceptual information and the dependence on interplay between direct and representational information. I examine how landmarks constitute a visual aid to navigation and how auditory aids present a ready alternative. I also investigate the ways in which team-based tasks afford new perspective and change the dynamics of a traditional navigation task.

Perceptual and symbolic cues

Perceptual cues hold many advantages over symbolic and linguistic methods. Unlike arbitrary symbols, perception functions as a signal that is fast and generally accurate (Neisser, 1978) and not reliant on linguistic skill or understanding (culturally neutral). However, the transmission of perceptual information has important limitations, making it less than optimal or even dangerous in the wrong setting. Perceptual signals apply only to those things that immediately surround us. In the case of highway driving, for example, perception keeps the car on the road, but the route between the driver and the ultimate destination is not necessarily available.

The simplicity that makes perception so useful does not negate the need for symbolic representation and reasoning with perceptual cues. Higher order cognition and symbolic problem solving can overcome such obstacles. Imagine standing on one side of a raging river with a pot of gold on the other. Perception relays information quickly and efficiently on the position of your goal, but this information is not enough to reach it (Reitman, 1965; Newell and Simon, 1972). Problem solving identifies the logs and boulders as a potential makeshift bridge to cross the river.

Interplay of Perceptual and Symbolic Cues

Represented information can compliment perceptual information. Problem solvers understand and anticipate threats, goals, obstacles, and any other relevant information an unlimited amount of time and space removed from the stimulus itself. Given the advantages of each of source of information, a balance of perceptual and symbolic cues could create more efficient performance. Poulton (1950) demonstrated the advantages of cognitive priming on traditionally perceptual tasks. Participants were instructed to respond as quickly as possible to every instance of an auditory cue. In one condition participants were informed there would be a second cue. In the other condition the second cue came without warning (i.e. without priming). As one might expect, the primed group performed significantly better. A similar effect can be demonstrated by a GPS system in a car. Drivers will less likely miss a turn when primed to look for it.

In a study of non-vocal auditory warning signals, Guillaume, Pellieux, Chastres, and Drake (2003) found that the cognitive framework of alarms impacted the perceived urgency. In other words there is a representational, interpretive aspect of alarms. Simulated alarms have acoustic characteristics that cause them to be perceived as more or less urgent, but the researchers determined that perceptual qualities alone were not sufficient in creating an optimal system. Sequence structure and the associated interpretation proved just as important, indicating that even if the perceptual cues are intact, a violation of how the participant thinks the alarm should work diminishes response likelihood. The findings in these experiments clearly demonstrated the benefits of utilizing cognitive and linguistic capabilities for otherwise perceptual tasks.

Landmarks

Landmarks in an environment can provide a type of cognitive facilitator of navigation. While they support perceptual engagement, they also require interpretation with regard to the environment. Landmarks are distinct features of an environment that serve as a point of reference. They are often large and visible from long distances in multiple directions. Research on spatial cognition has demonstrated that mental models of large spaces tend to be more qualitative than quantitative (e.g., Hirtle & Jonidas, 1985; Holyoak & Mah, 1982; Tversky, 1981), suggesting that landmarks may represent a more congruous way of relaying spatial information than measurements or headings.

Daniel and Denis (2004) asked participants to give route directions across a familiar environment and found that, when told to be as concise as possible, landmarks corresponding to actions (e.g., "Turn at the statue") were the least dispensable instructions. The finding implies that a prominent feature of a terrain present at a time of decision or action is most salient for relaying navigational information.

It is possible, however, that real-time navigation tasks (as opposed to recalled or planned tasks) rely less on landmarks. Lee and Tversky (2005) hint at this by noting that it is the encoding of landmarks and their relation to the environment, rather than landmarks in and of themselves, that facilitate performance. This suggests the necessity of a priori knowledge of the environment for landmarks to be useful. If a man were placed in Washington D.C. with no prior knowledge of the city layout and told to go to the White House, it would make little difference to him if he were placed at the Capitol building or a nondescript street corner. Environments could also be devoid of distinguishing characteristics usable as landmarks to the uninitiated, e.g., a jungle or desert. These issues also raise the question of what constitutes a landmark. Perhaps the physical characteristics are incidental, and the true definition of a landmark depends entirely on the cognitive constructions of the operator.

Hypothesis 1: Cognitive and perceptual aids will make navigation more efficient.

Auditory Aids

Audition can provide both direct and representational information, similar to vision, but does not directly interfere with visual processing, which may already be heavily engaged during navigation tasks. When the situation allows for users to wear headphones and sufficiently accurate tracking equipment, audio software currently available can spatialize a sound source to create a“3D” auditory perception. In this way, a software application can essentially create an aural reference point with limited cognitive demand on the receiver. Indeed, a sound source could update continuously and guide the receiver to a destination.

Gilkey et al. (2007) demonstrated that this type of audio display led to faster navigation times than did a handheld visual navigation display representing the same information. The already high visual demand of navigation tasks may have contributed to the audio display’s superiority. Also, visual displays require symbolic interpretation (or at least a scaling calculation) whereas spatialized audio is entirely dependent on perceptual information.

Coordinated Navigation with Teams

Investigating a team navigation task has both methodological and practical merit. Methodologically, team tasks provide a naturalistic language process measure while navigating. Separating the members of the team ensures that both members participate in the navigation as opposed to one leader taking charge. The rendezvous requirement forces communication while simultaneously reflecting the demands of a realistic military context. Limiting proximity to a shared visually accessible landmark makes the task more difficult, with greater reliance on real-time coordination.

Common ground. Measuring coordination is important in identifying team navigation techniques. Researchers have identified linguistic features as a means of analyzing degree of coordination (e.g. Clark, 1996; Clark, Schreuder, and Buttrick, 1983; Clark and Wilkes-Gibbs, 1986; Kramer, Oh, and Fussell, 2006; Wilkes-Gibbs and Clark, 1992). Clark, Schreuder, and Buttrick (1983) established that demonstrative language (the, this, etc.) requires common ground between speakers for understanding. As understanding is a primary component of coordinated speech, use of demonstrative language implies that the speaker believes he has common ground with the listener. Pronouns (he, I, they, etc.) also denote a certain degree of understanding between speakers.

Hypothesis 2: Language behavior will differ in response to the presence of cognitive and perceptual aids.

Hypothesis 2a: Use of pronouns will increase in the presence of cognitive and perceptual aids.

Junk Talk. Auditory cues could reduce navigation difficulty. Hancock and Meshkati (1988) described a method for measuring spare cognitive capacity called the auxiliary tasks method, where participants engage in non-essential tasks to varying degrees. This effect could manifest in team-based tasks (such as the current study) by a language content shift from task specific to generic.

Hypothesis 2b: Off-task conversation will increase in the presence of cognitive and perceptual aids.

Frame of reference. Frame of reference may also change as a function of shared understanding (Newcombe and Huttenlocher, 2000). Alternative frames of reference include spatial (representing objects as they related to one another, including the self) or absolute and abstract (utilizing the cardinal directions). Spatial language indicates the participants are navigating using their immediate surroundings, eschewing more abstract representations in favor of more accessible perceptual information.

Hypothesis 2c: Use of spatial frame of reference language will increase in the presence of cognitive and perceptual aids.

Hypothesis 2d: Use of absolute frame of reference language will decrease in the presence of cognitive and perceptual aids.

Process Account for a Team Navigation Performance Task

Hampton et al. (2012) found that separated teams using spatialized (push-to-talk) audio technology were able to rendezvous faster in unfamiliar environments than those with standard communication channels. In these types of environments each user hears his partner's voice as if it were coming from the direction in which his partner is standing, relative to himself. The researchers also manipulated the presence of added landmarks in the environments but did not find a significant difference in rendezvous time. The lack of a landmark effect introduces the idea that spatialized audio can interact with, or even supplant the traditional role played by landmarks. The current study examines the process data from the Hampton et al. experiment.

The advantage in spatialized audio conditions is promising, but the researchers did not explore the mechanism by which increased efficiency arose. Alternatively, strategies employed by different teams may have interacted significantly with the audio and landmark conditions. The present analysis aims to investigate the processes that resulted in the observed pattern of performance using linguistic measures. The inclusion of spatialized audio creates a new function of auditory communication. Participants can now use talk as direct *and* representational information simultaneously through the automatic creation of continuously updated spatial information based on the location of the source. The analysis of language will investigate how teams utilize this ability.

Whereas previous experimentation tested equivalent information presented to different sensory systems, I intend to test what happens when auditory displays are used in conjunction with visible landmarks. As audio pathways demonstrated a potential superiority in the research noted above, they may well supersede traditional landmark navigation.

II. Method

I obtained information on linguistic coordination and references from transcripts of an experiment already conducted by myself and a team of psychologists and computer scientists at Wright State University, the Air Force Research Laboratory, and DaytaOhio.

Participants

A total of eight paid participants worked in teams of two, with two all male teams, one all female team, and one male-female team. One member of each team played the role of a pararescue jumper and the other member played the role of a downed pilot to be rescued. The participants ranged in age from 21 to 29 years old. All participants were from the participant panel of the Battlespace Acoustics Branch at Wright-Patterson AFB and had normal hearing and corrected to normal vision. We dismissed one participant after a single practice trial because of nausea. We brought in a new participant to take her place, restarting the trials for that team.

Design

All pairs experienced all 60 treatment combinations in a 2 (audio conditions) * 2 (landmark conditions) * 15 (trials) repeated measures design. The order of the four combinations of conditions (Monaural/Added Landmark, Monaural/No Added Landmark, Spatialized Audio/Added Landmark, and Spatialized Audio/No Added Landmark) was partially counterbalanced across the four teams as follows. Added Landmark/No Added Landmark conditions were run 15 trials at a time within a given audio condition so that both landmark conditions would be experienced before switching to the second audio condition, where the landmark conditions would be repeated in the same order**.** We varied the order in which teams experienced conditions by having half the teams experience spatialized audio first, then mono and the other two teams the reverse. One team in each of those conditions experienced Added Landmark trials first and the other team experienced No Added Landmark trials first. As a result, none of the four teams had exactly the same order of treatment combinations (see Table 1)

Table 1

Visual representation of counterbalancing

Team 1		**Team 2**		**Team 3**		**Team 4**	
Mono	***ALM***	Spatial	NALM	Spatial	***ALM***	Mono	NALM
	NALM		***ALM***		NALM		***ALM***
Spatial	***ALM***	Mono	NALM	Mono	***ALM***	Spatial	NALM
	NALM		***ALM***		NALM		***ALM***

Note. Order of testing proceeded from top to bottom for each team. ***ALM*** stands for "added landmark" and NALM stands for "no added landmark".

Apparatus and manipulation

We conducted the experiment using two similar facilities. The Wright State Virtual Environment Research, Interactive Technology, And Simulation (VERITAS) facility at Wright-Patterson AFB contains CAVE®, a room-sized, five projection-surface (four walls and a floor) virtual environment display system. High-resolution (1200x1200 pixel) stereoscopic images are rendered with Barco Galaxy NW-12 DLP projectors via RealD CrystalEyes shutter glasses. An Intersense IS-900 tracking system monitors the position of the head and handheld Wand. Wright State University's Appenzeller Visualization Laboratory (AVL), operated by Wright State Applied Research Corporation, contains a room-sized, four projection-surface (three walls and a floor) virtual environment display system (I-Space, Barco). High-resolution (1400x1050 pixel) stereoscopic images are rendered with Barco Galaxy NW-6 Classic+ DLP projectors via RealD CrystalEyes shutter glasses. An optical tracking system (ARRTTRACK) monitors the position of the head and handheld wand. VERITAS and AVL are connected via an Internet2 wide area network (WAN) connection. We used the Distributed Interactive Simulation (DIS) messaging standard to communicate over this network, with a DIS software router to send local DIS messages across the wide area network. In both facilities, sounds were spatialized (with individualized head related transfer functions) using slab3d (v6.5.0; Miller and Wenzel, 2002) and presented via Sennheiser HMD-280-XQ headsets. Close-talking microphones on the headsets allowed participants in the two facilities to talk to each other. We used the DIS radio protocol to transmit voice communications over the network.

To “move” in the VERITAS environment the participant pointed the wand and pushed the joystick on the wand in the desired direction of travel. Because the display system in the AVL does not include a rear projection surface, it is not viable to move in that direction (i.e., the participant could not see where their avatar was going. Therefore, we used foot pedals in the AVL that allowed participants to rotate the virtual environment around them. That is, instead of turning to face the back wall and pointing the wand/joystick in that desired direction of travel as a participant in the VERITAS might do, the participant in the AVL would turn the environment with the foot pedals so that the imagery that would have been projected on the back wall is projected on the front (or a side) wall. Participants learned this system quickly, allowing them to move through the entire virtual environment and providing a substantially immersive experience. In both environments, a direction indicator appeared in the center of the base of all walls displaying primary and secondary compass coordinates (N, NW, W, SW, etc.) that changed with head position to let the participants know what direction they were facing.

In the monaural audio display condition, the DIS radio functioned much like a conventional push-to-talk radio and was actuated by a button on the participants' wand. In the spatialized audio display condition, the communications were spatialized so that they were heard as arising from the environment around the listener. Under the control of the talker (based on which of two buttons were pressed on the wand), the listener would hear the talker's voice as coming from the talker's direction or from the direction of an object in the environment that the talker marked using the wand (the audio annotation capability). To implement audio annotation, the talker pressed the appropriate button on the wand and utilized an already visible virtual laser to designate a target in the environment. This action caused the program to create a reference point on the designated spot from which the talker's voice would appear to emanate relative to the position of the listener. For example, a participant's avatar could be walking down the street, see a distinct building in the distance, point the wand at this building while pressing the correct button, and the partner would perceive the voice as arising from that building. So that it was clear to the listener which communication capability the talker was using, we superimposed a series of four broadband chirps on the communication channel immediately after the audio annotation capability was activated. Under the No Added Landmark condition, no structures were taller than five stories. Under the Added Landmark condition, additional structures (landmarks) were added to the same 15 terrains (two per terrain spatially separated and on different "tiles", which will be explained in the next paragraph). We placed the additional landmarks on spaces that were unoccupied in the No Added Landmark condition.

Task environments

Each trial took place within one of 15 virtual terrains, each measuring 500m x 500m. We constructed each terrain from four of six 250m x 250m “tiles.” The tiles could be put together so that any edge in one could be matched up with any edge of another, regardless of orientation. A large number of terrains could be created using different tiles in different orientations. We created the tiles using Presagis Creator Pro software in the Openflight file format. We used each of the six tiles an equal number of times in constructing the 15 terrains. Tile location and orientation of the tiles was not systematically controlled. Each tile contained various possible travel paths, from wide streets to narrow alleyways. Tiles included varying numbers and types of structures, ranging from windowless one-story sheds to large five-story structures with many windows. Colors of buildings also varied widely.

The tiles had commercial and residential areas modeled after a generic Moroccan city, whereas the landmarks (see Figures 1 & 2) reflected architecturally and culturally distinct styles (e.g., an Indian sculpture, a modern clock tower, a conventional American water tower, etc.). The landmarks were also designed to be taller than any other structure in the terrains, so that they could be seen from a distance and potentially be used as points of reference by the team members. The landmarks, when present, were added to previously vacant locations in the terrains so that no structures were deleted in creating the Added Landmark version of a terrain from the No Added Landmark version. In each terrain, two pairs of starting locations were selected by arbitrarily placing one participant near one edge of the terrain (not necessarily within line-of-sight of the edge) and then placing the other participant on the other side of the terrain at roughly the opposite longitude and latitude. The participants were assigned to the two positions (one for the Pararescue jumper and one for the downed pilot); reversing the assignments allowed the reuse of terrains across experimental conditions.

Figures 1 and 2. Examples of landmarks used in the ALM condition (Hampton et al. (2012)).

Nausea screening

This test consisted of both a self-report and a physical segment. Participants completed the self-report measure (Appendix A) first, responding to items concerning the participant's physical wellbeing both pre- and post-data collection. Items concerned level of nausea, comfort, and clarity of vision. Once completed and assessed, researchers instructed participants to stand on one leg with eyes closed and arms folded for 30 seconds. Then the researchers told the participants to switch legs and perform the test again. Researchers counted the number of adjustments (putting down the leg held in the air). Researchers repeated both procedures after participants completed testing for the day and compared the results to ensure that the testing had not caused any physical distress. If the researcher noticed a decrement in condition, he instructed to participant to sit down for at least five minutes and offered the participant water. The only time this occurred was in the one participant who experienced such levels of discomfort as to be dismissed from further participation.

Measures

Inter-rater reliability. Although psychological studies involving manual transcription do not generally assess inter-rater reliability for this process, I was concerned that my inherent interest in specific characteristics of the language presented could unconsciously bias my perceptions. I therefore conducted inter-rater reliability on the transcripts themselves as well as one measure that was not easily machine-readable.

I performed the analysis on the practice trials which I had conducted before data collection. Practice trials included every team and treatment combination. In this way I guarded against contaminating experimental data with an untested transcription process. As such, I had 27 practice trial audio files which could be tested without impacting the relevant experiment files.

The experiment audio program recorded every transmission to an audio file. I extracted metadata from these files indicating relative transmission time as well as experiment-specific data such as trial number and origin of transmission (i.e. which participant was talking). Ialso broke the audio file into component files corresponding to individual transmissions. Using another program I was able to sort these files by trial time (indexed at the end of the transmission) and add transcripts that were automatically integrated into the file metadata. Once I had recorded all of the transmissions in a trial, I extracted the metadata to a text file and coded it with the name of the team, trial, and audio condition.

Transcription. I recruited a recent psychology graduate to generate a second transcription process in exchange for a letter of recommendation based on her work for me. She was provided with written instructions regarding the transcription process (Appendix E). The duration of the rater's transcription process lasted roughly six hours total.

I then compared the two sets of transcripts. In order to determine how similar our transcripts were, I discarded superficial differences (e.g. "gonna" vs. "gunna") but counted any substantive deviations (e.g. "uh" vs. "a") as a function of total word count. I counted phrased discrepancies (a phrase found in one transcript but not the other) as individual words. I calculated the final score simply as total discrepancies over total word count and found agreement of just over 95%.

Junk Talk. I did not constrain participant language in the experiment and therefore participants tended to go off topic from time to time. Off-topic conversation, hereafter referred to as junk talk, includes any conversation substantively unrelated to the task with which the participants were presented, and used either for amusement or to provide a perceptual feedback signal. This presents two issues. One is that the off-topic conversation could influence all other linguistic features in a way unrelated to the task or experimental conditions. The second is that junk talk indicates a change in either strategy or task demand. In order to measure junk talk by automated scoring, it would be necessary to define by heuristic either junk talk (a prohibitively large linguistic space), or the task-relevant conversation, which assumes considerable knowledge *a priori*. I considered both options too susceptible to measurement error and therefore had to extract junk talk manually.

I also measured inter-rater reliability regarding junk talk classification, again on the practice trials and with the same secondary rater as the transcripts. In order to compound potential measurement errors and thus create a more stringent criterion for assessing inter-rater reliability, I instructed the secondary rater to extract junk talk from her own transcripts rather than mine. Instructions for this task (Appendix F) also included a brief explanation of the nature of the task to give context to the conversation. I did not disclose dependent measures (other than junk talk), so as to avoid biasing her ratings. Based on all 27 practice trials, the secondary rater and I had similar judgments, with the secondary rater scoring 205 words in junk talk to my 217. Aside from three instances of a word at the beginning or end of a phrase being judged differently, the disparity came entirely from one phrase that the secondary rater missed, concerning work for a class both participants were taking. Inter-rater reliability on classification of junk talk had a Cohen's kappa coefficient of .956, characterized by Landis & Koch (1977) as almost perfect agreement.

After establishing acceptable inter-rater reliability, I proceeded to extract junk talk from the test trials using the same identification procedure. I input the number of extracted words from each trial into an .xml spreadsheet and also kept the phrases intact in a .doc file in case further linguistic analyses into the junk talk were necessary. All remaining variables are scored without these phrases.

Computer measured variables. I created a java script program designed to measure various linguistic features of conversation (listed below) when presented with

.txt files. I input the transcripts, one file corresponding to every trial for a total of 240 files.

Landmark references. References to added landmarks provide a simple numerical value representing reliance on cognitive aids. To measure this, I created a dictionary of terms relating to the eight additional landmarks. In order to ensure completeness, I first listened to the trials of each team and noted the first reference to each landmark. Often teams generated unique descriptions that I could not have predicted *a priori* (e.g. the Indian-inspired archway was described as "the arch with the trash bags on top" and thereafter simply as "the trash bags"). Once I had finished the dictionary, I created a class within the same program I had written to count all references across teams.

Frame of reference. In order to measure the perceived frame of reference of a speaker, I divided the possible outcomes into either spatial or absolute. For absolute words, I counted all uses of the cardinal directions and combinations thereof. For spatial words, I used the spatial word category of the Linguistic Inquiry Word Count (LIWC; Pennebaker, Booth, & Francis, 2001) dictionary, minus cardinal directions.

Word count. My script automatically tallied total word count per trial, perhaps the most direct measure of communication efficiency. Words, non-fluencies, and fillers counted towards the trial total word count, which included all transmissions from the time participants were first positioned in starting positions until rendezvous.

Transmissions per trial. Number of transmissions per trial could provide a metric for understanding between teammates, with more back-and-forth potentially indicating adjustments to heading or position. Transmissions include any time participants engage the push-to-talk button and make an utterance.

Pronouns. A count of pronouns used per trial reflects shared understanding of referents. The list of pronouns once again comes from the Linguistic Inquiry Word Count 2001 dictionary.

Articles. Definite articles indicate shared understanding of referents whereas indefinite articles do not. A simple count of the use of “the,” “a,” and “an” per trial may demonstrate a certain kind of coordination.

Junk talk. Separately calculating junk talk provides a safeguard against non-task related conversation influencing the rest of the measures, as well as illustrating coping of the participants with the task (i.e., perhaps they have spare cognitive capacity if they are chatting). Junk talk includes any conversation not relating to the defined task. Parsing out junk talk required me to do a preliminary examination of the transcripts in order to develop identifying heuristics. I calculated junk talk by word count rather than transmissions, eliminating complications from long or complex transmissions. Frame of reference, article, and pronoun counts within the junk talk did not count towards those totals.

Task

The participants were told to imagine themselves in the roles of either a downed pilot stranded somewhere in a city or a pararescue jumper attempting to rescue the downed pilot. We instructed participants that the environment was finite and that if they began a simulation near a border of the map, his/her partner would likely start near the opposite border. The only communication available to the participants was the radio system controlled by their wands. Participants were encouraged to let one another know where they were in the environment and what they were doing in order to coordinate a rendezvous as quickly as possible. When the Pararescue jumper and downed pilot came within three meters of one another, a message appeared on the front screen indicating that the trial was complete. Trials were terminated if the team had failed to rendezvous in ten minutes. The rendezvous time was recorded as ten minutes in these cases.

Each terrain contained a total of eight systematically deployed hostile computer entities (enemy soldiers). Two hostiles appeared near the starting position and moved toward that position via a route that would take approximately one minute. Two additional, stationary hostiles blocked the most obvious or direct route between the Pararescue jumper and downed pilot. In the Added Landmark condition, one stationary hostile was placed at each landmark. The remaining stationary hostiles (two in the Added Landmark conditions, four in the No Added Landmark conditions) were placed in locations near likely travel paths for one or both participants. The hostiles were programmed to shoot the participants on sight. The hostiles had an effective firing range of exactly 100 meters, and when they fired they did not miss. Participants were instructed to "do your absolute best to avoid getting shot." When shot, the participants would hear a gunshot sound, which was presented with spatialized audio in all conditions, allowing the participant to localize the threat aurally. A burst of blue pixels emanated from the participant's location, similar to a small firework. There was no penalty for being shot. Participants continued the trial irrespective of number of times shot.

The pararescue jumper was armed with a “virtual gun” actuated by a button on the wand, but was instructed not to fire on hostiles unless fired upon or if he/she could see the downed pilot and the hostile was blocking the path to the downed pilot. The pararescue jumper had no limit on the range of his/her weapon except for the limitations of screen resolution. The downed pilot was unarmed. Terrains also hosted a varying number of civilian entities, but all had at least 20. The pararescue jumper was instructed never to shoot civilians.

Procedure

Before beginning data collection each day for each participant, researchers at both locations administered a pre-simulator balance and nausea test to participants (Appendix A). Then data collection began. One team of two participants performed at a time. The experimenter read the instructions for the task of either Pararescue jumper or downed pilot (Appendices B and C) depending on the participant’s assigned role as well as instructions for the appropriate audio condition and answered any questions from the participants. One team, after extra training, seemed to have trouble understanding the fundamental structure of the task and controls. I drafted additional instructions (Appendix D) that researchers read to both team members once. After the participants received task instructions, they each put on a vest that contained the wireless audio receiver, headphones with the head tracker attached to the headband, and shutter glasses. The participants sat in chairs and held the wand during the trials.

The experiment included four training trials for each of the audio conditions, two that contained landmarks and two that did not. These trials were comparable to actual trials except that the teammates were placed much closer together and encountered a higher concentration of hostiles so that they would experience being shot and more easily recognize that event. On the first day of either audio condition, teams experienced two training trials consecutively and were given an additional chance to ask questions before data collection began. On subsequent days under the same condition, only one training trial was conducted before data collection began. The experiment required between 8 and 10 days of participation from each team. IRB limitations required that participants spend no more than 30 minutes in the simulator continuously and no more than one hour per day. We therefore divided one hour sessions into two roughly 30 minute blocks of data collection.

III. Results

A series of repeated measures ANOVAs using audio, landmark manipulation, and trial tested the hypotheses for different sets of dependent measures. The dependent measures included a set of raw language measures, raw measures scaled as a percentage of overall word count without junk talk, and percentage measures scaled by task completion time. Data collected in this paper concern strategy and coordination. Hampton et al. (2012) collected performance data on the same trials previously, and a summary can be found in Appendix G. Given the small number of teams, I used an error term aggregated from all interactions involving team variability. This is consistent with the analysis of the performance data.

Only 2 of the 240 data points were missing entirely and 1 was missing a completion time. To reduce the resulting problem of confounding with subjects, I ran all relevant ANOVAs twice: once as an unbalanced design (leaving blank data blank) and once using mean substitution. I calculated mean substitution values by taking the grand mean plus the differences between the grand mean and the mean for each appropriate level of block, landmark manipulation, and audio main effects. For variables calculated as a function of word count and completion time, I used the calculated linguistic measure divided by the calculated word count and completion time instead of using direct mean substitution. All significant effects reported below reflect the higher p value between the unbalanced design and the mean substitution design. Mean values are from the mean substitution design.

Table 2

Summary of ANOVAs with raw language measures

Linguistic Variable	Audio	Audio*Landmark	Landmark
Indefinite Articles	XX		
Definite Articles			
Pronouns	XX		
Filler Words	XX		
Landmark References	XX	XX	XX
Spatial References	XX		
Absolute References	XX		
Junk Talk	XX		
Transmissions		X	
Word Count	XX		

Note. X indicates significance at the $p < 0.05$ level. XX indicates significance at the $p < 0.01$ level.

Raw Measures

As summarized in Table 2, when not adjusted for word count or response time, transmissions showed an interaction of audio and landmark manipulations. Landmark references also showed an interaction, but given the inherent dependency on the landmark manipulation, the interpretation of this effect is ambiguous. Transmissions increased in the presence of added landmarks with spatialized audio, but decreased without spatialized audio (see Figure 3). The simple effect for audio condition was not significant in the Added Landmarks condition (F (1, 42) = .65, p = .42) but was significant in the No Added Landmark condition (F (1, 42) = 26.63, p < .01).

The audio manipulation significantly impacted all linguistic variables with the exception of definite articles and transmissions. All means, including those that were not significantly different, were smaller in the spatial audio condition (see Table 3).

Table 3

Descriptive summary of significant effects for raw language measures by audio manipulation

Linguistic Variable	**Measure**	**Mono audio**	**Spatialized audio**
Indefinite Articles	Mean	8.49	4.46
	SD	7.71	4.66
Pronouns	Mean	31.46	19.81
	SD	20.27	14.89
Filler Words	Mean	12.48	6.84
	SD	13.57	7.79
Landmark references	Mean	1.62	0.83
	SD	3.10	1.60
Spatial references	Mean	29.71	14.89
	SD	23.68	12.78
Absolute references	Mean	8.51	3.46
	SD	5.68	3.64
Junk Talk	Mean	46.62	10.53
	SD	77.43	32.51
Word Count	Mean	351.91	198.34
	SD	236.42	144.95

Note. Values denote average words in each category per trial per team.

This illustrates the need for analyses based on word count and response time, as the established performance advantage could diminish linguistic variables regardless of strategy or perceived mutual understanding (i.e. there is less time to talk). The only significant impact of the landmark manipulation was on landmark references, which increased in the presence of added landmarks from M = .05, SD = .32 to M = 2.38, SD = 3.10.

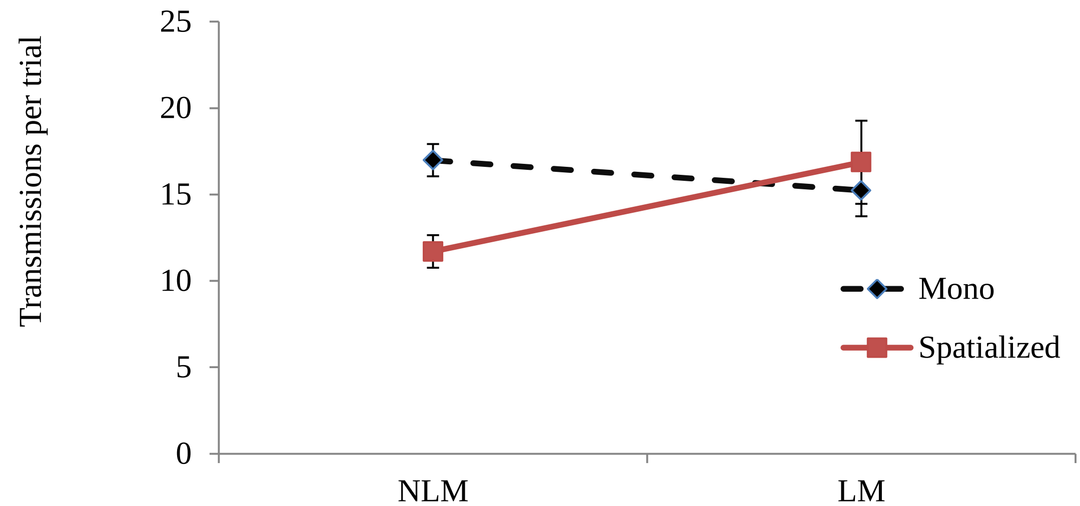

Figure 3. Interaction direction for transmissions. Transmissions per trial are on the Y axis with landmark condition on the X axis. The separate lines reflect audio condition. Error

Table 4

Summary of ANOVAs adjusted for word count

Linguistic Variable	Audio	Audio*Landmark	Landmark
Indefinite Articles/WC		XX	X
Definite Articles/WC			
Pronouns/WC	XX	XX	
Filler Words/WC			X
Landmark References/WC			XX
Spatial References/WC	XX	X	X
Absolute References/WC			
Junk Talk			
Transmissions			

Note. X indicates significance at the $p < 0.05$ level. XX indicates significance at the $p < 0.01$ level.

Measures Adjusted for Word Count

Adjusted for word count (see Table 4), indefinite articles, pronouns, and spatial frame of reference showed an interaction between audio and landmark manipulations. When landmarks were present, indefinite articles decreased slightly in the mono condition (from $M = .023$, $SD = .013$ to $M = .021$, $SD = .010$), but increased in the spatialized audio condition (from $M = .015$, $SD = .012$ to $M = .025$, $SD = .014$, see Figure 4). The simple effect for audio condition is not significant in the Added Landmark condition (F (1, 42) = .2.31, $p = .13$) but is significant in the No Added Landmark condition (F (1, 42) = 19.36, $p < .01$).

Pronouns increased when landmarks were present in the mono condition (from $M = .089$, $SD = .021$ to $M = .095$, $SD = .023$), but decreased with spatialized audio when landmarks were present (from $M = .111$, $SD = .035$ to $M = .094$, $SD = .029$, see Figure 5). The simple effect for audio condition is not significant in the Added Landmark condition (F (1, 42) = .05, $p = .83$) but is significant in the No Added Landmark condition (F (1, 42) = 23.75, $p < .01$).

Spatial references were roughly the same across landmark conditions in the mono audio condition (without landmarks $M = .082$, $SD = .022$ and with landmarks $M = .081$, $SD = .022$), but increased in the spatialized audio condition when landmarks were present (from $M = .061$, $SD = .028$ to $M = .078$, $SD = .024$, see Figure 6). The simple effect for audio condition is not significant in the Added Landmark condition (F (1, 42) = .49, $p = .49$) but is significant in the No Added Landmark condition (F (1, 42) = 28.93, $p < .01$).

Audio condition significantly impacted pronouns and spatial references. Pronouns increased in the spatialized audio condition (from $M = .092$, $SD = .022$ to $M = .103$, $SD = .033$), but spatial references decreased (from $M = .081$, $SD = .022$ to $M = .061$, $SD = .027$).

The addition of landmarks significantly increased the percentage of conversation that was indefinite articles (from $M = .019$, $SD = .013$ to $M = .023$, $SD = .012$), filler words (from *M* = .028, *SD* = .020 to *M* = .033, *SD* = .021), and spatial references (from *M* = .072, *SD* = .027 to *M* = .080, *SD* = .023).

Table 5

Descriptive summary of significant effects for language measures over word count in landmark manipulation

Linguistic Variable	**Measure**	**NALM**	**ALM**
Indefinite Articles	Mean	.019	.023
	SD	.013	.012
Filler Words	Mean	.028	.033
	SD	.020	.021
Landmark references	Mean	.000	.008
	SD	.001	.010
Spatial references	Mean	.072	.080
	SD	.027	.023

Note. NALM stands for "No Added Landmark" and ALM stands for "Added Landmark". Values denote percentage of conversation (e.g. .019 = 1.9% of conversation).

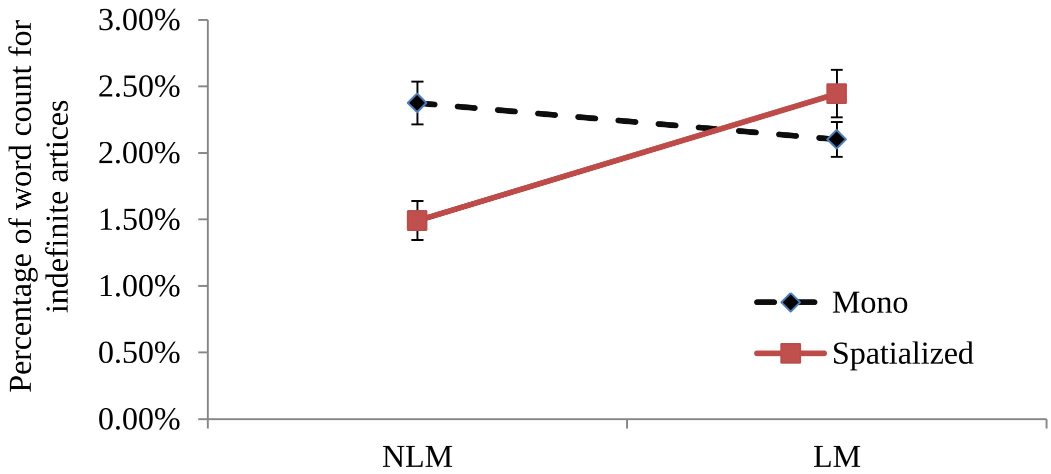

Figure 4. Direction of interaction for indefinite articles adjusted for word count.

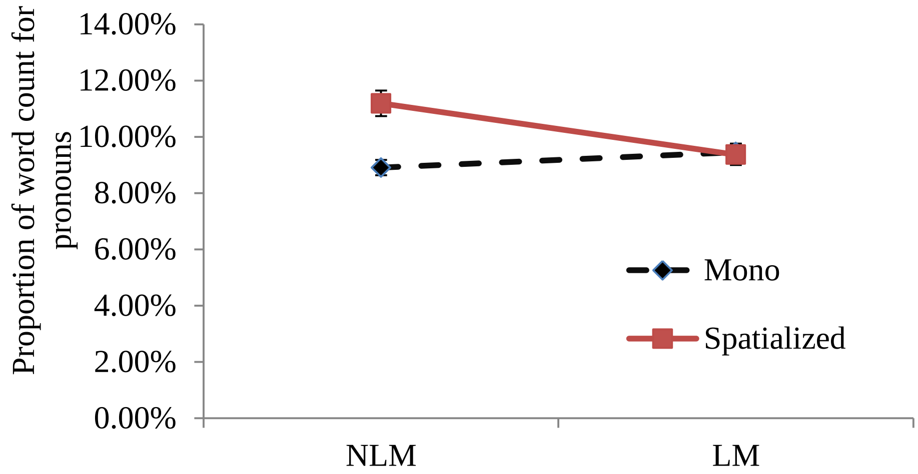

Figure 5. Direction of interaction for pronouns adjusted for word count.

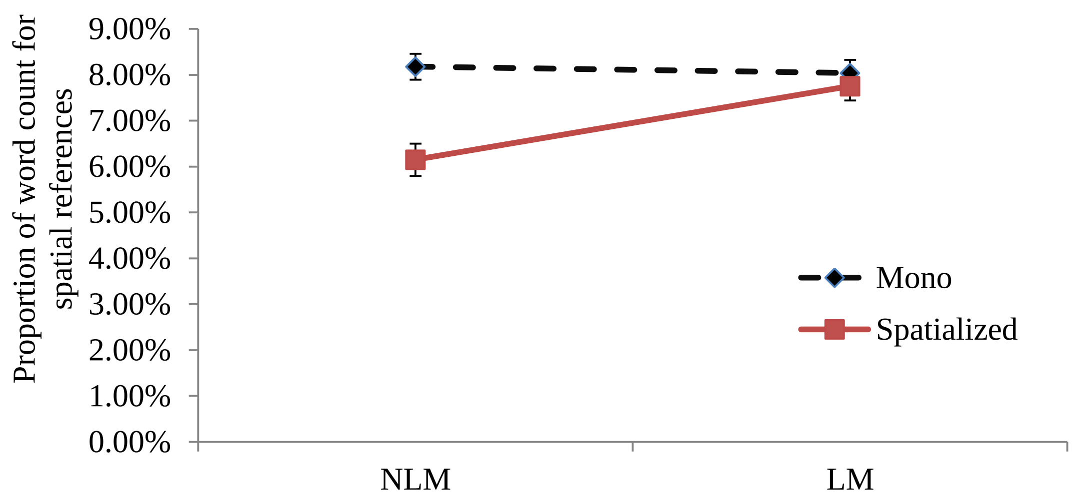

Figure 6. Direction of interaction for spatial references adjusted for word count.

Table 6

Summary of ANOVAs adjusted for completion time

Linguistic Variable	Audio	Audio*Landmark	Landmark
Indefinite Articles/WC	XX		
Definite Articles/WC			
Pronouns/WC	XX		
Filler Words/WC	XX		
Landmark References/WC	XX	XX	XX
Spatial References/WC	XX		
Absolute References/WC	XX		
Junk Talk	XX		
Transmissions	XX	X	
Word Count	XX		

Note. X indicates significance at the $p < 0.05$ level. XX indicates significance at the $p < 0.01$ level.

Measures Adjusted for Completion Time

Adjusted for completion time (yielding percentage of trial time devoted to a particular variable), transmissions showed an interaction between audio and landmark manipulations (see Figures 7). Landmark references also showed an interaction, but as with the raw measures, the interpretation is distinct from other manipulations due to its reliance on the landmark manipulation. In both cases, however, the direction of the interaction was similar to the direction displayed with unadjusted data. Spatialized audio had a significant impact on all linguistic variables except definite articles. All means decreased in the presence of spatialized audio, but this is largely a function of word count, which decreased from $M = 2.311$, $SD = 1.641$ to $M = 1.048$, $SD = .941$. This means that participants talked less than half as much of the time with spatialized audio available.

Presence of added landmarks significantly impacted only landmark references, which increased in the presence of added landmarks (from *M* = .000, *SD* = .001 to *M* = .015, *SD* = .020).

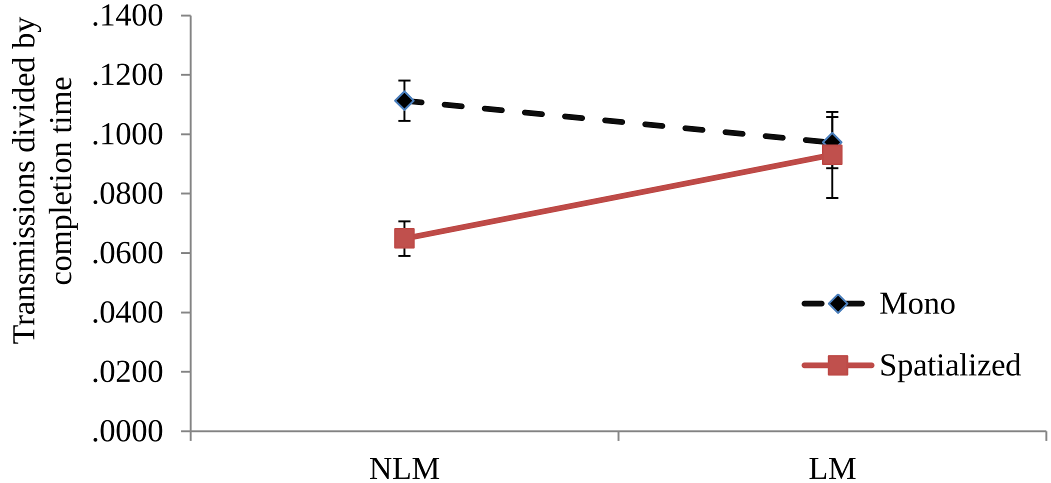

Figure 7. Direction of interaction for transmissions adjusted for completion time.

Raw Measures adjusted by word count and completion time

Table 7

Summary of significant effects across all levels of adjustment

Linguistic Variable	Audio	Audio*Landmark	Landmark
Indefinite Articles			
Definite Articles			
Pronouns	XX		
Filler Words			
Landmark References			XX
Spatial References	XX		
Absolute References			
Junk Talk			
Transmissions			
Word Count	XX		

Note. X indicates significance at the $p < 0.05$ level. XX indicates significance at the $p < 0.01$ level. Significant effects represent the highest p value measured across all levels of adjustment.

Table 7 illustrates which variables were significant in all adjustments (raw/adjusted for word count/adjusted for completion time). This represents the most conservative analysis including overall use, percentage of conversation, and percentage of trial time. No significant interactions withstand both adjustments. Also, every landmark manipulation effect except for references to landmarks disappears. Pronouns, spatial references, and word count all remain significant at the $p < .01$ level.

Dependent Measures as a Function of Performance

In order to test if it is reasonable to treat conversation density as a predictor of performance, I created a generalized linear model with word count over response time predicting response time. I used word count because it constitutes an aggregate of all other linguistic measures (except junk talk). Word count over response time measures density of conversation rather than total conversation (i.e., how often they talked, not how much). Total conversation would be biased by response time and thus not separable from response time for predictive validity. I also separated the results by team to avoid compounding error based on large team differences in both variables. The response time was positively skewed when plotted as a function of word count over response time, necessitating a square root correction of the variance. After this correction, three of the teams showed significant negative correlations at the $p < .05$ level (Team 1 $r = -.52$, $t = -4.68$, $p < .001$, Team 2 $r = -.47$, $t = -4.09$, $p < .001$, Team 4 $r = -.57$, $t = -5.22$, $p < .001$) and one team did not (Team 3 $r = -.24$, $t = -1.88$, $p = .066$).

Direct Quotations

Table 8 contains examples of types of phrases typically observed under various conditions. These phrases illustrate specific instances that contributed to overall statistical findings. Another interesting occurrence not captured by the linguistic measures was the use of non-verbal communication in the spatialized audio condition. Two of the teams explicitly stated strategies wherein they would ask their partner to simply make a noise that was spatially defined. Team 1 used “beat boxing” (free-form wordless rhythmic noises) and Team 3 would repeatedly click the audio annotation button wherein each transmission was preceded by spatialized “chirps.” This was the only use by any team of the audio annotation function outside of the practice trials. As such there are no linguistic data to analyze.

Table 8

Sample quotations by experimental condition

Quote #	Condition	Quote
1	spatialized audio	You're still sounding like you're right in front of me so I'm just gonna keep walking this way. Do I sound like I'm right in front of you? lalalalalalala.
2	spatialized audio	Just beat box, I'll constantly hear ya.
3	spatialized audio	It's kinda of irrelevant what's around us now, long as I know I'm walking right to you
4	spatialized audio	Alright I'm close to the east wall I'm facing north. Alright uh I'm gonna start heading towards the west wall. I think you're almost right in front of me now based on your voice so if I just, actually may- it might make it faster if you turned east and you start heading towards me too,
5	spatialized audio	You sound like you're kinda southward.
6	Mono audio	Ok when I face north I see, that's the edge of the map. What do you see when you see north? ok well I'm gonna head directly south you head directly north I think we got it. I'm near a pair of four big brown buildings with a skinny skinny white top. Any of that going on in yours?
7	Landmark	I don't maybe what you're calling a water tower is what I'm calling a fountain. Uh, no or kinda.
8	Mono audio	I'll help you one day whenever school lets me. I… That's all I see is this park and the palace. Yeah head there that's that's what I see. Are they green and orange? Yeah I see tons of them so you I'm surprised you don't see one of the palace steeples. Really. Ok so you're heading east. Ok. Yeah mike bought it he played… there you are.

IV. Discussion

My hypotheses concern the effects of cognitive and perceptual aids to navigation on linguistic measures. Landmarks and spatialized audio constitute cognitive and perceptual aids, respectively, and the linguistic measures reveal coordination and navigation strategy. Overall, the results seem to indicate a higher level of coordination in the presence of spatialized audio. The most striking difference between audio conditions is the greatly reduced word count, indicating that there was less need to communicate and presumably lower task demands. Participants explicitly stated that they knew where they needed to go without reference to visible features of the map. The inclusion of landmarks had no measureable impact on performance (Hampton, 2012), and seems to have interacted with the spatialized audio channel. Added landmarks only positively impacted coordination measures when spatialized audio was not available.

Cognitive and Perceptual Aids

Hypothesis 1 stated that cognitive and perceptual aids would make navigation more efficient. The data support this hypothesis in relation to perceptual aids (spatialized audio), but not cognitive (landmarks). Adjusting for completion time, all linguistic measures significantly decreased in use (except for definite articles, which showed no significant difference), indicating less need to explicitly coordinate in order to complete the task.

Interactions. The dominant pattern of findings is an interaction between the cognitive (landmark) aids and the perceptual (audio) aids. For example, an interaction appears between landmark and spatialized audio conditions where the highest percentage of conversation devoted to pronouns is in the absence of landmarks and the presence of spatialized audio (see Figure 6). The percentage is lower in the presence of added landmarks, approximately the same for either audio condition. The lowest percentage is with neither landmarks nor spatialized audio. In this way, any aid seems to augment pronoun use, but assuming that a higher percentage of pronouns reflects better mutual understanding, the best condition is with spatialized audio but without landmarks.

An interaction appears for spatial references (see Figure 7) similar to the interaction for pronouns. There is little change between landmark conditions within standard audio. Both means for spatialized audio are lower, but spatial references increase in the presence of added landmarks. This indicates a sort of navigational redundancy where participants use local navigation techniques (as opposed to a holistic perspective) because the environment affords their use, but without a significant performance advantage (Hampton, 2012). The transcripts seem to show that spatialized audio allows participants to see the environment on a larger scale, less dependent on the immediate surroundings. In Table 5, quotes 3 and 4 illustrate the shift in strategy when spatialized audio is available. Participants perceive one another's relative position but explicitly minimize the importance of terrain features. However, when landmarks are present they represent something noteworthy, and accordingly participants note them aloud, often using spatial words in their descriptions.

Landmarks. Despite the predominance of an interaction here, the established landmark advantages in the literature require consideration of landmarks as a main effect. The only stable effect of the presence of added landmarks was an increased reference to those landmarks. Adjusting for word count, indefinite articles and filler words both increased when added landmarks were available for reference. Because indefinite articles correspond to new information (Chafe, 1968), these two findings suggest lower levels of mutual understanding.

There are several possible reasons why I did not find a significant impact of landmarks on navigation despite the body of research suggesting otherwise. The dependence on teamwork may have played a part. It may be that landmarks are only useful for dyads when both team members can simultaneously see them. While the added landmarks in this test were large, there were many buildings impeding line-of-sight from an observer on the ground. This also introduces the possibility of physical characteristics of the landmarks used here being insufficient for the purposes of coordinated navigation.

One aspect where this paper differs from other landmark research is the task structure. It is possible that landmarks are useful for describing what one did or what one should do, but both require an overall knowledge of the environment. As Lee & Tversky (2005) suggested, it is not the landmarks themselves, but an accurate encoding of landmarks and the spatial relationships between them that is useful for navigation. Also landmarks have generally been studied as part of a stable environment. The current study asked each participant to find a moving target (namely, the other participant). The static landmarks may simply be extraneous in a dynamic task, or simply require too much coordination to be worthwhile in this particular task.

Another possibility is that this study did not actually include any true landmarks. It may be that landmarks are simply a construct defined by the situation, and their physical features only make them more or less likely to be used as a landmark.

Language and Team Navigation Processes

The above conclusions depend on hypotheses about the manner in which language reflects team navigation processes. Here, I compare the results to previously stated hypotheses regarding this relationship. In general, the measures did not behave entirely as hypothesized. In some cases, they were sensitive to one manipulation but not the other. In other cases, they showed no effect. And finally, in some cases, they behaved opposite to my hypotheses.

Manipulation sensitivity. Hypothesis 2a stated that the use of pronouns will increase in the presence of cognitive and perceptual aids. The data suggest that this hypothesis is mistaken in assuming that cognitive and perceptual aids will affect language patterns in the same way. Pronoun use did increase in the presence of spatialized audio, but there was no main effect for the landmark manipulation.

Hypotheses 2c and 2d stated that the presence of cognitive and perceptual aids will cause spatial language to increase and absolute frame of reference language to decrease, respectively. Recall that spatial language concerns objects in their spatial relations to one another whereas absolute frame of reference uses cardinal directions (North, South, East, and West). Once again, spatialized audio and landmarks cannot be taken as equivalent. Spatial references decreased in the presence of spatialized audio, but increased in the presence of added landmarks, indicating a fundamentally different strategy afforded by the different conditions.

Absolute frame of reference showed no significant difference across treatment combinations after adjusting for word count. Quotes 4 and 5, both with spatialized audio, specifically focus on absolute frame of reference without mentioning anything around them. Quote 6 illustrates the beginning of a mono audio trial where absolute frame of reference is used to orient both team members, but once general directions are established they shift to spatial relationships of surrounding features. This could mean that strategies regarding absolute frame of reference are different between audio conditions but they are not captured by percentage of conversation devoted to absolute frame of reference words.

The results for transmissions are particularly suspect. It was immediately obvious in observing the trials that teams, and individuals within teams, had vastly different strategies for when to transmit. Specifically, certain participants left the communication channel open for entire trial periods despite only speaking a small portion of the time. This created a transmission score of one for that trial, irrespective of linguistic strategy or coordination. Analyses based on average transmission length would likewise be compromised due to the influence of completion time.

Contradicting language behavior. Hypothesis 2b stated that off-topic conversation would increase in the presence of cognitive and perceptual aids to navigation. The data indicate only that junk talk decreased in the presence of spatialized audio (adjusted for completion time), meaning that in spatialized audio conditions, participants spent less relative time talking about things not related to the task. Again we see a distinction between cognitive and perceptual aids beyond what I predicted, but importantly the direction of the relationship between spatialized audio and junk talk is the opposite of my hypothesis. The increased junk talk in the mono audio condition is possibly due to the low gain between decision making and results. Participants could decide to go in a certain direction and while en route make conversation. Once participants established that they did not share perceptual data with one another, it could take some time before it would be reasonable to check again. Alternatively, participants may have used casual conversation as a means of reassuring their teammate of the status quo (i.e., letting them know that nothing has changed). Quote 8 in Table 5 illustrates a team discussing a videogame, then checking their surroundings, establishing they do not share a visual referent, and going back to discussing the game until they find one another. Teams may feel the need to fill silences when there is uncertainty attached. The spatialized audio would mean that participants always shared some perceptual information and therefore did not need to fill time.

Prediction of Completion Time

While time to rendezvous is not the only relevant aspect (type of strategy and perceived difficulty are also worthwhile), it does present the clearest indicator of performance. Three of the four teams demonstrated significant predictive validity of linguistic measures on completion time, with the fourth team marginally significant. This indicates that it is reasonable to treat linguistic measures as predictors of performance.

Taken on its own, the direction of the relationship indicates that density of conversation increases as completion time decreases, or, stated inversely, silence is bad for coordinated navigation. In a rendezvous task like the one simulated, shared perceptual referents are more likely to arise as participants become more proximal to one another. This gives participants more useful data worthy of conversation. At first glance the relationship seems contrary to the findings that spatialized audio lowered both conversation density and response time, but the correlation was averaged over both audio conditions indicating that the findings are especially true without spatialized audio. The failure of Team 3 to reach significance on this measure could be a function of strategic change particularly useful for spatialized audio. As stated earlier, Team 3 often used the spatialized "chirps" that accompanied transmissions without actually engaging in verbal exchange, creating an unmeasured substitute for conversation on navigation.

Limitations

Sample size is a chief concern for the validity of the study. With only eight total participants, organized into four teams, the results are likely to be somewhat unreliable, especially considering the unbounded variations expected in linguistic productions. Also, the teams all consisted of native English speakers in their twenties from the Dayton area, potentially introducing a cultural bias.

I instructed teams to avoid being shot and to only shoot enemies when necessary. Because there were no penalties, I introduced variability based on willingness to follow effectively arbitrary rules. Also, the nature of the simulation, specifically the similarity to common video game dynamics, gave an advantage to participants with experience playing video games that they would not necessarily have in a real world navigation task.

While I believe I have presented reasonable explanations for the lack of a landmark effect, the experiment manipulated both task structure and the presence of spatialized audio relative to prior landmark research. In other words, previous landmark studies did not include spatialized audio or large scale real-time navigation tasks in unfamiliar environments, let alone both together. The data from trials without spatialized audio suggest limitations of landmarks relative to task structure, but with only half of a limited data set, conclusions are difficult to draw.

Practical Applications

Open radio communication could be dangerous in enemy territory. With spatialized audio, even intercepted communications would be useless without a priori knowledge of combatants' ground positions. Non-verbal communication is possibly just as useful as standard radio communications (as demonstrated by Team 1's use of "beat boxing" to present continuously updated spatial information, and Team 3's continual use of the spatialized chirps preceding audio annotation transmissions). Using this technique, troops in combat scenarios could effectively organize themselves via open radio channels using nothing more than clicks that are completely indecipherable to enemies. Non-military applications are also possible in any environment in which computers and headphones are usable. As the experiment was essentially a computer simulation, video game applications are easy to imagine.

V. Conclusions

This paper expands the research areas of spatialized audio capabilities, as well as the use of landmarks with respect to large scale, real-time team navigation. Through the use of two integrated, immersive virtual reality simulators, I was able to create a controlled situation for testing how spatially separated teammates would interact with navigational aids, both technological and environmental, while moving toward a dynamic goal. Through linguistic analysis I investigated the performance advantage demonstrated by spatialized audio and the lack of a performance advantage demonstrated with additional landmarks (Hampton et al. 2012). This analysis indicated that these different types of aids (cognitive vs. perceptual) affect communication strategy (in terms of frame of reference and what information needed to be communicated) and not just performance, with result patterns consistent across measurement adjustments (for completion time and word count).

The most striking result from the linguistic analysis regarding spatialized audio is the overall decreased need for verbal communication. Word counts overall as well as nine of the ten individual linguistic measures decreased significantly (definite articles showed no significant change) in the presence of spatialized audio. Two of the teams demonstrated an explicit strategy of bypassing verbal communication altogether and relying on noise with no semantic content. Spatialized audio also correlated with a decrease in spatial frame of reference, indicating that teams were less dependent on their immediate surroundings.

The most striking result for the effect of added landmarks is largely the lack of any effect. Despite considerable experimental grounding, adding landmarks showed little linguistic impact on spontaneous team navigation. While there are a number of possible explanations for this, including physical characteristics of the landmarks used and the dynamic task structure, I believe the real issue is that landmarks are a purely cognitive construct (rather than being objectively defined, as we attempted), and thus only useful when understood in relation to the other aspects of the environment.

While sample size is clearly a concern for this study, the findings indicate that spatialized audio represents a considerable technological advancement to, at least, navigation. Teams demonstrate improved efficiency and mutual understanding, solidifying previous findings of performance advantage. This technology and the strategies emergent from its use truly have the potential to increase mission effectiveness and save lives in a variety of situations.

VI. Acknowledgements

This paper would not have been possible without the guidance and support of my advisor, Valerie Shalin, Ph.D. who kept me pointed in the right direction despite my many attempts to deviate. My thesis committee of Herb Colle, Ph.D. and Robert Gilkey, Ph.D. likewise offered valuable insights into both my experimental and writing techniques. Thanks also go to Dr. Gilkey, Brian Simpson Ph.D., and Victor Finnomore for letting me tag along on the project that I would eventually expand to create this study. Patient technical support from the computer magicians Jeffrey Cowgill, John Stewart, and Brian Lenahan immeasurably expedited the data analysis process. My thanks to Tessa Miracle and Michael Hartman who greatly aided the creation of my reliability measure for the transcript process and to Gary Burns, Ph.D. and David LaHuis, Ph.D. for guidance on statistical analysis and interpretation. Finally, thanks to my wife, Kate, who accepts my strange working hours and supports my ambition without reservation.

APPENDIX A

ID_________ Date__/__/__ PRE or POST

SIMULATOR SICKNESS QUESTIONNAIRE

Instructions: Circle the items that apply to you RIGHT NOW.

SYMPTOM		*RATING*		
1. General Discomfort	None	Slight	Moderate	Severe
2. Fatigue	None	Slight	Moderate	Severe
3. Headache	None	Slight	Moderate	Severe
4. Eye Strain	None	Slight	Moderate	Severe
5. Difficulty Focusing	None	Slight	Moderate	Severe
6. Increased Salivation	None	Slight	Moderate	Severe
7. Sweating	None	Slight	Moderate	Severe
8. Nausea	None	Slight	Moderate	Severe
9. Difficulty Concentrating	None	Slight	Moderate	Severe
10. "Fullness of the Head"	None	Slight	Moderate	Severe
11. Blurred Vision	None	Slight	Moderate	Severe
12. Dizzy (eyes open)	None	Slight	Moderate	Severe
13. Dizzy (eyes closed)	None	Slight	Moderate	Severe
14. Vertigo	None	Slight	Moderate	Severe
15. Stomach Awareness**	None	Slight	Moderate	Severe
16. Burping	None	Slight	Moderate	Severe

17. Other. Please describe____
____
____
____

** "Stomach Awareness" is usually used to indicate a feeling of discomfort which is just short of nausea.

APPENDIX B

Task instructions -Joshi

<u>Pararescue jumper</u>

You are a soldier trying to find a downed pilot in a city. Your task is to meet with the pilot. You will start in one location in the city, and the pilot will be somewhere else in the city. You will have to coordinate with the pilot in order to meet. You will have a radio to communicate with the pilot.

There are multiple enemies in the city. You will have a weapon, but the pilot will not. The rules of engagement state that you may fire only when fired upon or when you are within visual range of the pilot. If you are shot, a burst of blue will come from your virtual body and a gunshot sound will be audible. Do you absolute best not to get shot. You will be given a map of the city to help you. The pilot will not have a map so you may want to use prominent landmarks to coordinate. Be aware that the software will lock up every time a new scenario is loading. Do you have any questions on the task so far?

Wand/comms instructions

<u>Mono</u>

You will be using a standard radio to communicate with the other person. When you use the radio, they will hear your voice, but they will not have any information about your location or the location of enemies unless you tell them. Use the radio to tell your teammate where you are and what you’re doing. To talk to the other person, press and hold the button second from the left on your wand.

To shoot an enemy, point the laser at him and pull the trigger on the back of the wand.

Spatial

You will be using spatial audio to communicate with the other person. Use this to tell your teammate where you are and what you're doing. They will hear your voice as if it was coming from your location. To use the radio, press and hold the button second from the left on your wand. You will also be able to mark locations and make it sound as if your voice is coming from the place you are marking. To mark a location, point the laser at the location you wish to mark and hold the button second from the right on your wand. To shoot an enemy, point the laser at him and pull the trigger on the back of the wand.

Motion model

To move through the environment, move the joystick on the wand in the direction you would like to move. To change the direction you are facing, use the foot pedals. Pushing the right pedal forward will turn you to the left. Pushing the left pedal forward will turn you to the right. Keep in mind that the pedals are controlled by sliding, like a ski, rather than pivoting, like the accelerator in your car

APPENDIX C

Task Instructions- CAVE

Pilot

You are a pilot who has crashed in a city. Your task is to meet with a soldier who is trying to rescue you. You will start in one location in the city, and the soldier will be somewhere else in the city. You will have to coordinate with the soldier in order to meet. You will have a radio to communicate with the soldier, and the soldier will have an overhead, black and white map of the terrain.

There are multiple enemies in the city. You do not have a weapon, so you will have to avoid the enemies. Start moving quickly, because enemy soldiers will be moving towards your starting location. If you are shot, a burst of blue will come from your virtual body and a gunshot sound will be audible. Avoiding being shot is, of course, a high priority. Do your absolute best not to get shot.

Many of the terrains will have prominent landmarks that you may find useful to coordinate with the soldier. Be aware that the software will lock up every time a new scenario is loading.

Wand/comms instructions

Mono

You will be using a standard radio to communicate with the other person. They will hear your voice, but they will not have any information about your location or the location of enemies unless you tell them. Use the radio to tell your teammate where you are and what you're doing. To talk to the other person, press and hold the button second from the left on your wand.

Spatial

You will be using spatial audio to communicate with the other person. Use this to tell your teammate where you are and what you're doing. They will hear your voice as if it is coming from your location. To use the radio, press and hold the button second from the left on your wand. You will also be able to mark locations and make it sound as if your voice is coming from the place you are marking. To mark a location, point the laser at the location you wish to mark, then press and hold the button second from the right on your wand.

Motion Model

Use the joystick on your wand to move around the environment. The direction you select will be relative to where the laser is pointing. In other words, the direction the laser is pointing always corresponds to pressing forward on the joystick, not to the direction you yourself are facing. For example, you can change your heading by turning the wand without changing the joystick position. This also means you can physically turn in your chair without affecting your heading, as long as the wand stays still.

APPENDIX D

Additional Training Notes

Often the best thing to do is determine where you are on the map before proceeding. If you can see the edge of the map when you start, it's likely that your partner will be near the opposite edge. Coordinate with them using the direction indicators at the bottom of your screen (N, S, E, and W). Keep in mind that it shows which way your head is facing at any given time. And remember that the taller the building, the more likely your partner can also see it. No two terrains are exactly the same, so don't count on using the same landmarks in consecutive trials.

APPENDIX E

Notes for transcriber.

Getting started:

-On the desktop, open foobar2000. Bottom middle, with an alien face icon.

-Go to MY COMPUTER\veritas on 'veritas_master'(Z:)\task3_data\Transcripts by confirmation agent

-You will be writing transcripts for all of the .flac files, one trial at a time.

-Open a numbered folder. There should be folders inside labeled flac and wav. Click and drag the flac folder into the lower half of the foobar window. Multiple files should appear.

-Click the "track no" tab at the top of those files to put them in the correct order.

-Double click on the first file. The audio should begin playing. The audio will loop, meaning it will continue playing the same file over and over until you are finished transcribing.

-Hit the "t" key to open a transcription window. Click inside this window and transcribe the file.

-When finished with one file, click ok or hit "Enter." Then use the arrow keys to navigate to the next file. Press enter to play, and repeat. When you have finished all files in one trial, select all files in the foobar window, and hit "Delete." The transcripts are saved automatically. Navigate to the next available trial and repeat the entire process until all trials are transcribed.

Transcription key points:

-**DON'T** worry about CAPITALIZATION. Lowercase is fine for everything.

-DO worry about PUNCTUATION. Please know the difference between "it's" (it is) and "its". This is important for the analysis. Likewise "there", "their", "they're."

-Try to WRITE EXACTLY WHAT THEY SAY even if it's bad grammar and you know what they mean

(e.g. "I see a bunch of building" -Don't include an “s” if they don't)

-People will repeat words and include non-fluencies (uh, um, like) make sure you WRITE ALL OF THESE in the order they are said

-If non-fluencies (um, uh, etc.) are elongated, **DO NOT** write them differently (i.e. you should write “uh” regardless of how long they pronounced it)

-Use your best judgment on ambiguous sounds. If you are unsure, you **CANNOT** ask anyone else. That's the point of having both of us transcribe independently.

-Slang or informal wording should be transcribed as they sound (e.g. “gonna” for going to) but **DO** include full –ing endings to words even if the ‘g’ is not clearly pronounced. (e.g. “I’m coming toward you” instead of “I’m comin toward you”)

-Ignore laughter or throat clearing, but write down other kinds of non-language sounds with <brackets>.

-partially completed words should be indicated with a hyphen (e.g. “I s- I see where you are.”)

-Many transmissions will be only static. Ignore these and move on to the next one.

APPENDIX F

The study:
We had two participants in similar virtual reality facilities who were placed in the same virtual "city." They were placed in the roles of a downed pilot or a pararescue jumper. Their task was to rendezvous as quickly as possible, coordinating by radio. We're examining the effect of environmental cues and different radio systems on how they coordinate. One of the audio capabilities is "throwing" your voice or projecting it to a location removed from the self. Keep this in mind.

Instructions for quantifying transcription data:
Junk talk includes any conversation substantively unrelated to the task with which the participants were presented, and used either for amusement or to provide a perceptual feedback signal. Keep in mind that trials could have no junk talk.

Discussions on weekend plans or mutual acquaintances should be counted towards **junk talk**.
(ex. "We should get tacos later" is ***not*** relevant)
Discussions on what the simulation reminds them of, or how they feel about it, or strategy are fine. (ex. "We are really messing this trial up" is relevant)

When you encounter junk talk, copy and paste it into a Word document under the heading of the trial in which it occurs.

Any questions, don't hesitate to e-mail or call.
--****

APPENDIX G

Summary of the completion time data from the original study.

Treatment Combination	Mean	Standard Deviation
Mono Audio	216.59	107.57
Spatialized Audio	165.42	70.63
Additional Landmarks	185.88	75.07
No Additional Landmarks	196.13	110.42
Mono with Landmarks	214.36	92.46
Mono without Landmarks	218.82	121.58
Spatialized with Landmarks	157.40	34.17
Spatialized without Landmarks	173.44	93.61

Note. Data represent four teams each performing 60 trials in a 2x2 design.

VII. References

Bangerter, A. & Clark, H. (2003). Navigating joint projects with dialogue. *Cognitive Science, 27*, 195-225.

Bangerter, A., & Smolenski, C. (2000). *Die Wiederverwendung konzeptueller Pakte bei der Bezugnahme auf neue. Informationen.* [Reusing conceptual pacts in referring to new information]. Research report no. 72, Psychological Institute, University of Basel.

Clark, H.H. (1996). *Using Language.* Chicago, IL. Chicago University Press.

Clark, H. H., Schreuder, R., & Buttrick, S. (1983). Common ground and the understanding of demonstrative reference. *Journal of Verbal Learning and Verbal Behavior , 22*, 245- 258.

Clark, H. H., & Wilkes-Gibbs, D. (1986). Referring as a collaborative process. *Cognition , 22*, 1-39.

Chafe, WL. (1968). English noun inflection and related matters from a generative semantic point of view. Project on Linguistic Analysis Reports, 2nd Series, No. 6, 1-51

Dabbish, L., & Kraut, R. E. (2008). Awareness displays and social motivation for coordinating communication. *Information Systems Research*, 19(2), 221-238.

Daniel, M.P. & Denis, M. (2004). The production of route directions: Investigating conditions that favour conciseness is spatial discourse. *Applied Cognitive Psychology, 18*, 57-75.

Gilkey, R. H., Simpson, B. D., Brungart, D.L., Cowgill, J.L., & Ephram, A.J. (2007). 3D audio display for pararescue jumpers. Proceedings of the HFES 51st Annual Meeting, 51, 1349-1352.

Glaser, B.G. & Strauss, A.L. (1967). *The Discovery of grounded theory: Strategies for qualitative research.* Piscataway, NJ. Transaction Publishers.

Gibson, J. J. (1966). The senses considered as perceptual systems. Houghton-Mifflin, Boston.

Guillaume, A., Pellieux, L., Chastres, V., Drake, C. (2003). Judging the urgency of nonvocal auditory warning signals: Perceptual and cognitive processes. *Journal of Experimental Psychology: Applied*, 9(3), 196-212.

Hancock, P. & Meshkati, N. (1988). *Human mental workload.* New York, NY. Elsevier Science Publishing Company, Inc.

Hancock, P. & Szalma, J. (2008). Performance under stress. Orlando, FL. Ashgate Publishing. 115-120.

Hirtle, S. C. and Jonides, J. (1985). Evidence of hierarchies in cognitive maps. *Memory and*

Cognition, 13, 208-217.

Holyoak, K.J. & Mah, W. A. (1982). Cognitive reference points in judgments of symbolic magnitude. *Cognitive Psychology, 14*, 328-352.

Kramer, A.D., Oh, L.M., & Fussell, S.R. (2006). Using Linguistic Features to Measure Presence in Computer-Mediated Communication. *Human-Computer Interaction Institute. Paper 53*. http://repository.cmu.edu/hcii/53.

Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. Biometrics, 159-174.

Lee, P.U. & Tversky, B. (2005). Interplay between visual and spatial: The effect of landmark descriptions on comprehension of route/survey spatial descriptions. *Spatial Cognition and Computation. 5 (2&3),* 163-185.

Linguistic Inquiry and Word Count · James W. Pennebaker, Martha E. Francis, and Roger J. Booth. Lawerence Erlbaum Associates, Mahwah, NJ, (2001).

Mackworth, N. H. (1948). The breakdown of vigilance during prolonged visual search. *Quarterly Journal of Experimental Psychology*, 1, 6-21.

Neisser, U. (1976). Cognition and reality: Principles and implications of cognitive psychology. San Francisco: Freeman.

Neisser, U. (1978). Perceiving, anticipating and imagining. Minnesota Studies in the Philosophy of Science, 9, 89-106.

Newcombe, N. & Huttenlocher, J. (2000). Making space: The development of spatial representation and reasoning. Cambridge: MIT Press

Newell, Allen. (1990). *Unified theories of cognition.* Cambridge, MA. Harvard University Press.

Newell, A., & Simon, H. A. (1972). Human problem solving (Vol. 14). Englewood Cliffs, NJ: Prentice-Hall.

Ou, J., Oh, L. M., Fussell, S. R., Blum, T. & Yang, J. (2008). Predicting visual focus of attention from intention in remote collaborative tasks. *IEEE Transactions on Multimedia*.

Patten, C.J.D., Kircher, A., Östlund, J., & Nilsson, L. (2003). Using mobile telephones: Cognitive workload and attention resource allocation. *Accident Analysis and Prevention,* 36, 341-350.

Patten, C.J.D., Kircher, A., Östlund, J., Nilsson, L., & Svenson, O. (2005). Driver experience and cognitive workload in different traffic environments. *Accident Analysis and Prevention,* 38(5), 887-894.

Peirce, C. S., & Hoopes, J. E. (1991). Peirce on signs: Writings on semiotic. UNC Press Books.

Poulton, E.C. (1950). Perceptual anticipation and response time. *Quarterly Journal of Experimental Psychology,* 2(3), 99-112.

Reitman, W. R. (1965). Cognition and thought: An information-processing approach. New York: Wiley.

Smith, Bryan. (2003). The use of communication strategies in computer-mediated communication. *System*, *31*(1), 29-53.

Tversky, B. (1981). Distortions in memory for maps. *Cognitive Psychology, 13,* 407-433.

Tversky, B., Kim, J., & Cohen, A. (1981). Mental models of spatial relations and transformations from language. *Advance in Psychology*, *128,* 239-258.

Warm, J. (1977). The event rate context in vigilance: Relation to signal probability and expectancy. *Bulletin of the Psychonomic Society.*

Wolfe, J.M., Horowitz, T.S., Kenner, M., Hyle, N. & Vasan, N. (2004). How fast can you change your mind? The speed of top-down guidance in visual search. *Vision Research 44,* 1411-1426.